Individualized evaluation of cholinesterase inhibitors effects in dementia with adaptive cognitive testing

HANS WOUTERS,¹ JOS P.C.M. VAN CAMPEN,¹ BREGJE A. APPELS,² JOS H. BEIJNEN,³
AEILKO H. ZWINDERMAN,⁴ WILLEM A. VAN GOOL⁵ & BEN SCHMAND^{5,6}

¹ Department of Geriatric Medicine, Slotervaart Hospital, Amsterdam, The Netherlands

² Department of Medical Psychology and Hospital Psychiatry, Slotervaart Hospital, Amsterdam, The Netherlands

³ Department of Pharmacy & Pharmacology, Slotervaart Hospital, Amsterdam, The Netherlands

⁴ Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Academic Medical Centre, Amsterdam, The Netherlands

⁵ Department of Neurology, Academic Medical Centre, Amsterdam, The Netherlands

⁶ Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands

Key words

Alzheimer, Lewy Body, neuropsychology, psychometrics

Correspondence

Jos Van Campen, Department of Geriatric Medicine, Slotervaart Hospital, Louwesweg 6, 1066 EC Amsterdam, The Netherlands.
Telephone (+31) 20 512 93 33
Email: Jos.VanCampen@slz.nl

Received 1 October 2014;
revised 20 April 2015;
accepted 5 May 2015

Abstract

Computerized Adaptive Testing (CAT) of cognitive function, selects for every individual patient, only items of appropriate difficulty to estimate his or her level of cognitive impairment. Therefore, CAT has the potential to combine brevity with precision. We retrospectively examined the evaluation of treatment effects of cholinesterase inhibitors by CAT using longitudinal data from 643 patients from a Dutch teaching hospital who were diagnosed with Alzheimer disease or Lewy Body disease. The Cambridge Cognitive Examination (CAMCOG) was administered before treatment initiation and after intervals of six months of treatment. A previously validated CAT was simulated using 47 CAMCOG items. Results demonstrated that the CAT required a median number of 17 items (inter-quartile range 16–20), or a corresponding 64% test reduction, to estimate patients' global cognitive impairment levels. At the same time, intraclass correlations between global cognitive impairment levels as estimated by CAT or based on all 47 CAMCOG items, ranged from 0.93 at baseline to 0.91–0.94 at follow-up measurements. Slightly more people had substantial decline on the original CAMCOG ($N = 31/285$, 11%) than on the CAT ($N = 17/285$, 6%). We conclude that CAT saves time, does not lose much precision, and therefore deserves a role in the evaluation of treatment effects in dementia. Copyright © 2015 John Wiley & Sons, Ltd.

Introduction

Legal approval of pharmacological treatment of dementia by the European Medicines Agency and the US Food and Drug Administration requires demonstration of its efficacy by objective cognitive tests. Among the best known of these objective cognitive tests are the Cambridge Cognitive Examination (CAMCOG) (Roth *et al.*, 1986) in Europe and the cognitive part of the Alzheimer Disease Assessment Scale (ADAS-cog) (Rosen *et al.*, 1984) in the United States. The CAMCOG has been used frequently in population studies (Roth *et al.*, 1986; Cullum *et al.*, 2000) and the ADAS-cog in clinical trials (Raskind *et al.*, 2000). Important limitations of these instruments are that they are time-consuming, and also burdensome to patients. The time required for testing, usually about 30 to 45 minutes, make these instruments less practical for routine evaluation of the efficacy of (pharmacological) treatment in individual geriatric and neurologic outpatients in everyday practice. As a consequence, clinicians often opt for less precise but much briefer tests such as the Mini Mental Status Examination (MMSE) (Folstein *et al.*, 1975).

Item Response Theory (IRT), a modern psychometric theory, and the related Rasch measurement model offer promising solutions to this problem (Hobart *et al.*, 2007; Holman *et al.*, 2003). IRT models arrange the items of the test and patients on a “latent” common scale. Items are placed on the scale according to their level of difficulty, whereas patients are placed on the same scale according to their level of global cognitive impairment. More difficult items have a lower probability of a correct response even for patients whose scores on the latent scale correspond to relatively mild global cognitive impairment. Conversely, most patients tend to respond correctly to the easiest items, except for the patients who have severe global cognitive impairment. The scale is “latent” or only indirectly observable as it is statistically estimated from the patients’ observed responses to the items of the cognitive test. The items of tests like the CAMCOG, the ADAS-cog, and the MMSE measure cognitive functioning in various domains e.g. memory and language. Therefore, the latent scale is labelled “global cognitive impairment”, analogous to the label that would be assigned to a conventional total test score.

This “item difficulty-person impairment mapping” by IRT methods enables adaptive testing. Unlike the conventional total score approach that requires the administration of every item of the test to estimate patients’ global cognitive impairment levels, adaptive testing selects for each individual patient only items of appropriate difficulty. Too difficult and

too easy items are skipped. In practice, the decision whether or not to administer particular items is made by an algorithm that is continuously fed with each response by the patient during testing. This is called “Computerized Adaptive Testing” (CAT) (see Methods section). Because all items are calibrated on the same scale of global cognitive impairment, individuals’ impairment levels can still be compared even though different individuals were not given identical sets of items.

Adaptive testing has the potential to considerably shorten extensive cognitive tests and to make them much more patient-friendly and feasible for routine administration in outpatient settings. Our previous research (Wouters *et al.*, 2009a, 2009b, 2011) showed that adaptive testing indeed combines precision with feasibility. Compared to the whole test, test reductions of up to 40% to 60% were found with only slight loss of precision. Global cognitive impairment levels as estimated with adaptive testing were in excellent agreement with the estimates based on the whole test (intraclass correlations ranged from 0.95 to 0.99).

Despite its potential for clinical practice, adaptive cognitive testing has not yet been used to evaluate the efficacy of pharmacological treatment of dementia. Therefore, the purpose of this study was to report on applying adaptive cognitive testing to evaluate the efficacy of cholinesterase inhibitors in individual patients using the CAMCOG. We conducted retrospective analyses on longitudinal data from a memory clinic registry of patients with Alzheimer disease (AD) or Lewy Body disease (LBD) who were treated with rivastigmin or galantamin. Specifically, we examined the amount of test reduction of a previously developed CAT version of the CAMCOG as well as its agreement with a 47-item IRT validated CAMCOG and its correlation with the original CAMCOG. Furthermore, we examined whether rates of substantial cognitive decline differed between the original CAMCOG and the CAT version of the CAMCOG. The latter was examined, because substantial decline notwithstanding pharmacological treatment, was considered as a potential reason for treatment discontinuation.

Methods

Participants

Longitudinal data were retrieved from a registry of consecutive patients who were referred to the memory clinic of the Slotervaart Hospital in Amsterdam, the Netherlands from 2000 to 2009. These patients were subsequently treated with cholinesterase inhibitors, after being diagnosed with AD or

LBD. After excluding 13 patients (2%) who had > 10% missing responses on the CAMCOG due to patient related factors such as hearing deficits, impaired vision, insufficient command of the Dutch language or functional impairments of the dominant hand, data from 643 patients were available for analysis (see Figure 1, a flowchart, for further information).

At baseline, all patients had been diagnosed with dementia by means of a standardized dementia assessment that was consistent with international diagnostic consensus criteria (McKhann *et al.*, 1984; McKeith *et al.*, 1996). The standardized dementia assessment included a patient history, an informant interview, physical, neurologic, and psychiatric screenings, and a cognitive screening consisting of the MMSE and the Seven Minute Screen (Solomon *et al.*, 1998) as well as laboratory tests. On indication, a CT-scan or a magnetic resonance imaging scan of the brain and an extensive neuropsychological examination were performed.

To evaluate the efficacy of cholinesterase inhibitors, patients were regularly monitored at six-month intervals

using a standardized protocol. The protocol consisted of a cognitive examination using the CAMCOG, an assessment of behavioural problems with the Revised Memory and Behavioural Problems Checklist (RMBPC) (Teri *et al.*, 1992) and an assessment of the amount of assistance required to accomplish instrumental and basic activities of daily living using the Interview of Deterioration in Daily activities in Dementia (IDDD) (Teunisse and Derix, 1997). Trained test assistants administered the CAMCOG, patients' next of kin or caregivers filled out the RMBPC and the IDDD. The study was approved by the medical ethical committee of the Slotervaart Hospital.

Demographic and clinical characteristics

Demographic and clinical characteristics were extracted from patients' medical records. We scored educational level on a seven-point scale (1 = less than six years of primary school to 7 = university). Using the RMBPC items (all items scored on five-point scales: 0 never, 4 always), we calculated patients' memory complaints in daily life

Figure 1. Flowchart of patient inclusion and follow-up.

(items 1–7, maximum total score 28 points), their symptoms of depression (items 17–23, maximum total score 28 points) and disruptive behaviours (items 8–16, maximum total score 36 points). Patients' impairment to perform activities of daily living was calculated from the total score of the IDDD (11 five-point scale items; 0 never, 4 always, maximum total score 44 points). Global cognitive impairment was assessed with the total scores of the CAMCOG and the MMSE.

Computerized adaptive testing (CAT) algorithm

A previously validated CAT algorithm (Wouters *et al.*, 2009a, 2009b, 2011; Lindeboom *et al.*, 2004) was examined. Item difficulties of the CAT algorithm were fixed according to the item difficulties of a validated set of 47 CAMCOG items previously published (Lindeboom *et al.*, 2004). Lindeboom *et al.* (2004) applied specialized software to fit the One Parameter Logistic Model (OPLM), a Rasch related measurement model (Verhelst and Glas, 1995), using CAMCOG data from participants of a population study and patients from two memory clinics. In brief, the OPLM estimates a global cognitive impairment level for every patient based on his or her responding to every item and a difficulty level for each item based on the response of all patients. Both global cognitive impairment levels of patients and item difficulty levels are expressed on a common scale of global cognitive impairment. The scale is a log-odds scale that typically ranges from -2 to $+2$ and corresponds to the total score of the test. In the OPLM, the probability to respond correctly is an s-shaped function of the difference between the estimated patient's global cognitive impairment and the estimated item difficulty. By definition, the item difficulty corresponds to an impairment level at which the probability of responding correctly is 50%. Patients with more impairment than the item difficulty have a probability of a correct response that is lower than 50%, and patients with less impairment have a higher than 50% probability. OPLM validity analyses previously demonstrated that for 47 CAMCOG items valid difficulties could be estimated (Lindeboom *et al.*, 2004). These 47 CAMCOG items are henceforth labelled as the C-47. The remaining 13 items were not used in the CAT algorithm, but were retained to calculate the original CAMCOG total score.

Subsequently, the C-47 items were adaptively administered by the CAT algorithm, henceforth labelled as the *adaptive C-47*, to estimate each individual patient's global cognitive impairment level by only selecting items of appropriate difficulty. This was done in the way as described

elsewhere (Wouters *et al.*, 2009a, 2009b). In brief, a standard set of six initial starting items described elsewhere (Wouters *et al.*, 2009a) were always selected at the beginning of the adaptive C-47. This was done to ensure the selection of clinically important items, and to prevent possible administration of recall items prior to the registration items which would have resulted in a meaningless assessment of memory function.

After the selection of the initial item set, the CAT algorithm of the adaptive C-47 made a provisional estimate of a patient's global cognitive impairment level along with a standard error as a measure of reliability. Based on each subsequent response by the patient, the algorithm updated the global cognitive impairment estimate and the standard error by selecting a more difficult item in case of a correct answer and an easier item after an incorrect answer. With each response, the standard error decreased and the global cognitive impairment estimate became more reliable. The algorithm terminated after achieving a standard error of 0.15 log-odds units, reflecting ~80% reliability.

Statistical analyses

To examine the brevity of the adaptive C-47, we calculated the median reduction in number of items needed to generate a precise global cognitive impairment estimate and expressed this as a percentage relative to the number of items of the whole C-47 (or whole 47 CAMCOG item set) and the original 60-item CAMCOG. Brevity was examined separately for different measurement occasions. The agreement between the global cognitive impairment estimates generated with the adaptive C-47 and those based on the whole C-47 item set was examined with intraclass correlations (two-way-mixed). Associations between the impairment estimates by the adaptive C-47 and the total scores on the original CAMCOG were examined with Spearman's rank correlations. Agreement and associations were also examined separately for different measurement occasions. We also calculated Pearson's correlations to examine associations of cognitive change from baseline to the first follow-up either measured with the original CAMCOG or with the adaptive C-47, with change in memory problems in daily life and with change of assistance needed to accomplish activities of daily living from baseline to the first follow-up.

The adaptive C-47 was compared with the original CAMCOG with regard to its sensitivity for substantial cognitive decline. Substantial decline notwithstanding pharmacological treatment, was considered as a potential reason for treatment discontinuation. To prevent bias, this

analysis was restricted to the change scores between baseline and the first follow-up. Substantial cognitive decline in individual patients was defined as follows. First, we calculated standard errors of the estimates of global cognitive impairment from a historical cohort of untreated or control patients who were referred to the memory clinic of the Academic Medical Centre in Amsterdam ($n = 190$) described elsewhere (Walstra *et al.*, 1997). Item difficulties for the historical cohort were fixed for the adaptive C-47 according to the item difficulties of the C-47 (Lindeboom *et al.*, 2004) and for the original CAMCOG according to the item difficulties estimated with the data from the treated patients of the Slotervaart Hospital. The 95% confidence intervals were calculated as $\pm 1.96 \times$ the standard error of each global cognitive impairment level as estimated with the data from the control patients of the historical cohort. Second, we estimated the global cognitive impairment levels of the patients of our study or those from the registry of the Slotervaart Hospital, while keeping the item difficulties fixed. Subsequently, we plotted their estimated global cognitive impairment levels at baseline (horizontal axis) against their declines of estimated impairment or their impairment level at the first follow-up minus that at baseline (vertical axis). Decline was judged against the 95% confidence intervals as calculated with the data from the control patients, and was deemed substantial when exceeding the lower limit of the 95% confidence interval. Although for the original CAMCOG it would have been more intuitive to adopt the total score metric, we chose the OPLM log-odds impairment estimates for both the adaptive C-47 and the original CAMCOG in order to compare both in a common frame of reference.

Results

Baseline demographic and clinical characteristics of the study participants are shown in Table 1. Over two-thirds were women, patients had on average received intermediate education and notable spread was found for age. The majority of the patients were diagnosed with AD, approximately 15% had a diagnosis of AD with concomitant cerebrovascular disease, and approximately 10% were diagnosed with LBD. Mild to moderate levels of depressive symptoms, memory problems in daily life, disruptive behaviours and required assistance to accomplish instrumental and basic activities of daily living were observed. As expected, the CAMCOG and MMSE revealed considerable global cognitive impairment as the average scores fell well below the cutoff points for dementia (MMSE cutoff point 23/24, CAMCOG cutoff point 79/80).

Median test length reductions as achieved with the adaptive C-47 were 64% (inter-quartile range 57–66%) compared to the whole C-47 item set (the 47 CAMCOG item set) at baseline because the adaptive C-47 needed a median number of 17 items (inter-quartile range 16–20) to estimate reliable global cognitive impairment estimates. Compared to the whole original CAMCOG, this test reduction was 72% (inter-quartile range 67–73%). At follow-up measurement occasions, median test reductions compared to the whole C-47 item set varied from 62% to 64% and median test reductions compared to the whole original CAMCOG varied from 70% to 72%. Overall, no reduction was achieved for 1% of the assessments. Intraclass correlations between the adaptive and the whole C-47, ranged from 0.93 at baseline to 0.91–0.94 at any of the follow-ups. High correlations between the adaptive C-47 and the original CAMCOG were observed: Spearman's rank correlation = 0.83 at baseline and Spearman's rank correlations ranging from 0.83 to 0.86 at follow-up measurement occasions (see Figure 2 for graphical displays).

Significant but modest associations were observed between change from baseline to the first follow-up in memory problems in daily life as reported by patients' informants and cognitive change from baseline to the first follow-up as measured with either the original CAMCOG ($r = -0.25$, $p < 0.01$) or the adaptive C-47 ($r = -0.17$, $p < 0.01$). The association between informant reported change from baseline to the first follow-up in assistance needed to accomplish activities of daily living and cognitive change from baseline to the first follow-up was small regardless of whether it was measured with the original CAMCOG or the adaptive C-47 and only significant when cognitive decline was measured with the original CAMCOG (Spearman's rank correlations: original CAMCOG $r = -0.24$, $p < 0.01$ versus adaptive C-47 $r = -0.10$, $p > 0.05$).

Figure 3 shows, for the adaptive C-47 and the original CAMCOG, the change in estimated global cognitive impairment of patients (y -axis) plotted against their baseline global cognitive impairment levels (x -axis). On the group level, cognitive impairment remained quite stable irrespective of whether cognitive decline was measured with the original CAMCOG or adaptive C-47 (mean change adaptive C-47 0.005, standard deviation 0.21 versus mean change original CAMCOG 0.054, standard deviation 0.43). The 95% confidence intervals based on the estimated global cognitive impairment levels of the control patients showed that slightly less patients had substantial decline on the adaptive C-47 ($N = 17/285$, 6%) than on the original CAMCOG ($N = 31/285$, 11%).

Table 1. Baseline demographic and clinical characteristics (*N* = 643)

Variables	Statistic
<i>Demographic characteristics</i>	
Women (<i>n</i> , %)	447 (70)
Educational level (seven-point scale, mean, standard deviation)	3.7 (1.5)
Age (mean, standard deviation)	80.5 (6.0)
<i>Clinical characteristics</i>	
Dementia diagnosis (<i>n</i> , %)	
Alzheimer disease	497 (77)
Alzheimer disease with concomitant cerebrovascular disease	90 (14)
Lewy Body disease	56 (9)
Depressive symptoms RMBPC, maximum score 28 (median, interquartile range)	6 (2-11)
Disruptive behaviours RMBPC, maximum score 36 (median, interquartile range)	4 (2-7)
Activities of daily living IDDD, maximum score 44 (median, interquartile range)	14 (7-20)
<i>Cognitive function</i>	
Subjective memory complaints, maximum score 28 (median, interquartile range)	17 (14-19)
CAMCOG: global cognition (mean, standard deviation)	63.8 (12.9)
MMSE: global cognition (mean, standard deviation)	19.2 (4.3)

Note: RMBPC, Revised Memory and Behavioural Problems Checklist; IDDD, Interview of Deterioration in Daily activities in Dementia; CAMCOG, Cambridge Cognitive Examination; MMSE, Mini Mental Status Examination

Discussion

Consistent with previous findings (Wouters *et al.*, 2009a, 2009b, 2011), the results of this study support the potential of adaptive cognitive testing to establish an optimal trade-off between precision and brevity. A substantial test reduction of about 60% was observed for the adaptive C-47 compared to the whole C-47 (47 CAMCOG item set). The adaptive C-47 was even 70% shorter than the original CAMCOG. At the same time, the adaptive C-47 estimates of global cognitive impairment showed good agreement with scores calculated from the whole C-47 and were highly correlated with the original CAMCOG. Although these findings support the potential applicability of adaptive testing for the evaluation of treatment of cognitive impairment with pharmacotherapy, a possible caveat was the finding that a somewhat lower proportion of patients had decline on the adaptive C-47 than on the whole CAMCOG (6% versus 11%). This finding needs further examination as in the absence of a gold standard, the reason for this finding cannot be determined. Given the inconclusiveness of this finding, one can only speculate about the consequences of this difference. Adopting the adaptive C-47 instead of the CAMCOG in clinical practice could result in decisions to continue treatment with cholinesterase inhibitors for ~5% more patients. Given that cholinesterase inhibitors have bothersome side effects that occur frequently, the

consequence in turn for physicians would be an increased need to monitor patients' side effects. At the same time, dangerous side effects of cholinesterase inhibitors are rare. Hence adoption of adaptive cognitive testing is not likely to increase the harm done to patients. Ultimately, as long as for individual patients the benefits in terms of cognitive improvement outweigh the downsides in terms of bothersome side effects, continuation with treatment could be considered. Of course, CAT should never be used in isolation. Rather, it should form an integral part of a standardized protocol to evaluate the benefits of a therapeutic intervention. As described in the Methods section, such a standardized protocol should not only consist of a cognitive examination, but also an assessment of behavioural problems and activities of daily living as well as side effects.

The clinical relevance of adaptive testing is also evident. It is likely to reconcile the requirement of demonstrating the efficacy of pharmacological treatment using objective cognitive tests with a manner of cognitive testing that is sufficiently brief to make it feasible for the everyday practice of busy outpatient clinics. Also, it is less burdensome to patients. As a consequence, the relative brevity of the test might actually give a better impression of the cognition of the patients due to less fatigue. However, whether adaptive cognitive testing indeed reduces patients' fatigue and thereby improves the validity of cognitive testing needs further prospective examination.

Figure 2. Associations between adaptive C-47 estimates of global cognitive impairment with whole C-47 and original CAMCOG estimates of global cognitive impairment.

Figure 3. Six month change of estimated global cognitive impairment of treated patients (y axis) plotted against their baseline estimated global cognitive impairment (x axis) for the adaptive C-47 and the whole CAMCOG. Estimated global cognitive impairment levels are expressed in log-odds. Dotted lines represent 95% confidence intervals derived from standard errors of estimated global cognitive impairment levels of untreated control patients.

This study had several strengths. With regard to the data, the large sample size, the evaluation of the efficacy of cholinesterase inhibitors using a standardized protocol of a cognitive, a behavioural and an activities of daily living measure as well as the standardized dementia assessment constitute real strengths. Several limitations of this study that give rise to suggestions for further research are also worth mentioning. First, analyses of this paper were based on previously collected longitudinal data. Although we have previously conducted a prospective validation of adaptive testing (Wouters *et al.*, 2011), this prospective validation was a cross-sectional study aimed at examining whether adaptive testing enabled efficient and precise screening for dementia versus normal ageing. Prospective

longitudinal studies, e.g. clinical trials that adopt adaptive cognitive testing are therefore also worthwhile to pursue. Second, application of adaptive testing to other cognitive tests than the CAMCOG would be an avenue for further research. Third, because the present sample included only few patients with LBD, replications with greater samples of LBD patients are important. Fourth, as we defined substantial cognitive decline of our patients using historical data from untreated patients, replications of our findings using newly collected normative data from untreated patients, could be fruitful. A final general concern regarding adaptive testing is the skipping of clinically important items. We addressed this problem by administering a set of initial items at the beginning of the test. This also

prevented the potential administration of the recall items prior to the registration items by the adaptive testing algorithm, which would have resulted in a meaningless assessment of memory function.

There might be other study opportunities as well. First, the “item difficulty-person impairment mapping” by IRT methods not only allows for CAT. It also allows for the development of subtests with varying difficulty levels that are accommodated to varying impairment levels, or a kind of semi-adaptive testing. For example, administering a subtest with sensitive neuropsychological tests and difficult CAMCOG items such as the serial 7s to patients with Mild Cognitive Impairment or early AD and a subtest with easier CAMCOG items such as counting backwards from 20 to 1 and neuropsychological tests aimed at more severe cognitive impairment to patients with moderate to severe dementia. This might be particularly useful for cohort studies that follow people in time for many years. Second, development of adaptive testing algorithms for the grading of the severity of behavioural symptoms might also be fruitful to explore as was done already for activities of daily living (Weisscher *et al.*, 2007). And so may be adaptive testing algorithms meant for grading impairment in cognitive domains that are especially affected by AD such as memory,

perceptual speed and executive function (Bäckman *et al.*, 2005). Finally, a worthwhile area of research would be how adaptive cognitive testing relates to volumetric and functional brain measures analogous to previous research in which associations between conventional cognitive testing and brain parameters were studied (Van Der Flier *et al.*, 2002).

Conclusion

Taken together, our findings suggest that adaptive cognitive testing allows for an evaluation of the benefits of pharmacological treatment of dementia that is feasible, more patient-friendly and sufficiently precise.

Acknowledgement

The work presented in this paper was financially supported by the Department of Geriatric Medicine of the Slotervaart Hospital.

Declaration of interest statement

The authors have no competing interests.

References

- Bäckman L., Jones S., Berger A.K., Laukka E.J., Small B.J. (2005) Cognitive impairment in pre-clinical Alzheimer's disease: a meta-analysis. *Neuropsychology*, **19**, 520–531.
- Cullum S., Huppert F.A., McGee M., Denning T., Ahmed A., Paykel E.S., Brayne C. (2000) Decline across different domains of cognitive function in normal ageing: results of a longitudinal population-based study using CAMCOG. *International Journal of Geriatric Psychiatry*, **15**, 853–862.
- Folstein M.F., Folstein S.E., McHugh P.R. (1975) 'Mini-mental state' a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, **12**, 189–198.
- Hobart J.C., Cano S.J., Zajicek J.P., Thompson A.J. (2007) Rating scales as outcome measures for clinical trials in neurology: problems, solutions, and recommendations. *Lancet Neurology*, **6**, 1094–1105.
- Holman R., Lindeboom R., Glas C.A.W., Vermeulen M., De Haan R.J. (2003) Constructing an item bank using Item Response Theory: the AMC linear disability score project. *Health Services and Outcomes Research Methodology*, **4**, 19–33.
- Lindeboom R., Schmand B., Holman R., De Haan R.J., Vermeulen M. (2004) Improved brief assessment of cognition in aging and dementia. *Neurology*, **63**, 543–546.
- McKeith I.G., Galasko D., Kosaka K., Perry E.K., Dickson D.W., Hansen L.A., Salmon D.P., Lowe J., Mirra S.S., Byrne E.J., Lennox G., Quinn N.P., Edwardson J.A., Ince P.G., Bergeron C., Burns A., Miller B.L., Lovestone S., Collerton D., Jansen E.N., Ballard C., de Vos R.A., Wilcock G.K., Jellinger K.A., Perry R.H. (1996) Consensus guidelines for the clinical and pathologic diagnosis of dementia with Lewy bodies (DLB): report of the consortium on DLB international workshop. *Neurology*, **47**, 1113–1124.
- McKhann G., Drachman D., Folstein M., Katzman R., Price D., Stadlan E.M. (1984) Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, **34**, 939–944.
- Raskind M.A., Peskind E.R., Wessel T., Yuan W. (2000) Galantamine in AD: a 6-month randomized, placebo-controlled trial with a 6-month extension. The Galantamine USA-1 Study Group. *Neurology*, **54**, 2261–2268.
- Rosen W.G., Mohs R.C., Davis K.L. (1984) A new rating scale for Alzheimer's disease. *American Journal of Psychiatry*, **141**, 1356–1364.
- Roth M., Tym E., Mountjoy C.Q., Huppert F.A., Hendrie H., Verma S., Goddard R. CAMDEX. (1986) A standardized instrument for the diagnosis of mental disorder in the elderly with special reference to early detection of dementia. *British Journal of Psychiatry*, **149**, 698–709.
- Solomon P.R., Hirschhoff A., Kelly B., Relin M., Brush M., DeVeaux R.D., Pendlebury W.W. (1998) A 7 minute neurocognitive screening battery highly sensitive to Alzheimer's disease. *Archives of Neurology*, **55**, 349–355.
- Teri L., Truax P., Logsdon R., Uomoto J., Zarit S., Vitaliano P.P. (1992) Assessment of behavioral problems in dementia: the revised memory and behavior problems checklist. *Psychology and Aging*, **7**, 622–631.

- Teunisse S., Derix M.M. (1997) The interview for deterioration in daily living activities in dementia: agreement between primary and secondary caregivers. *International Psychogeriatrics*, **9**(S1), 155–162.
- Van Der Flier W.M., Van Den Heuvel D.M., Weverling-Rijnsburger A.W., Spilt A., Bollen E. L., Westendorp R.G., Middelkoop H.A., Van Buchem M.A. (2002) Cognitive decline in AD and mild cognitive impairment is associated with global brain damage. *Neurology*, **59**, 874–879.
- Verhelst N.D., Glas C.A.W. (1995) The one parameter logistic model. In Fischer G.H., Molenaar I.W. (eds) *Rasch Models, Foundations, Recent Developments and Applications*, pp. 215–237, New York, Springer Verlag.
- Walstra G.J., Teunisse S., Van Gool W.A., Van Crevel H. (1997) Reversible dementia in elderly patients referred to a memory clinic. *Journal of Neurology*, **244**, 17–22.
- Weisscher N., Post B., de Haan R.J., Glas C.A.W., Speelman J.D., Vermeulen M. (2007) The AMC Linear Disability Score in patients with newly diagnosed Parkinson disease. *Neurology*, **69**, 2155–2161.
- Wouters H., Zwiderman A.H., Van Gool W.A., Schmand B., Lindeboom R. (2009a) Adaptive cognitive testing in dementia. *International Journal of Methods in Psychiatric Research*, **18**, 118–127.
- Wouters H., de Koning I., Zwiderman A.H., Van Gool W.A., Schmand B., Buitter M., Lindeboom R. (2009b) Adaptive cognitive testing in cerebrovascular disease and vascular dementia. *Dementia and Geriatric Cognitive Disorders*, **28**, 486–492.
- Wouters H., Van Campen J., Appels B., Lindeboom R., Buitter M., De Haan R.J., Zwiderman A.H., Van Gool W.A., Schmand B. (2011) Does adaptive cognitive testing combine efficiency with precision? Prospective findings. *Journal of Alzheimers Disease*, **25**, 595–603.